



Departamento de Estadística
Universidad Carlos III de Madrid

BIOESTADISTICA (55 - 10536)

Evaluación de pruebas diagnósticas

CONCEPTOS CLAVE

- 1) Características del diseño en un estudio para evaluar pruebas diagnósticas.
- 2) Medidas estadísticas para evaluar pruebas diagnósticas: sensibilidad, especificidad, valores predictivos, y cociente de probabilidades.
- 3) Corrección de los valores predictivos a partir del Teorema de Bayes
- 4) Correcciones de las medidas de asociación en estudios de casos y controles, y en estudios de cohortes.

1. INTRODUCCION

Se trata, en este tema, de estudiar aquellos procedimientos estadísticos que permiten evaluar la validez, es decir cuánto se aproxima una medida al valor *real* que pretende medir. Son procedimientos muy generales, pero aquí se van a tratar al hilo del problema de las **pruebas diagnósticas** (problema muy importante en la práctica y la investigación clínica): pruebas para determinar si un individuo tiene, o no, una cierta enfermedad (u otra característica). Hay que resaltar (y no siempre se hace) que si bien la validez de una prueba depende exclusivamente de la prueba, las estimaciones numéricas que se obtengan de la misma pueden depender de la prevalencia y de las características clínicas de los pacientes que se hayan estudiado.

2. INDICES DE VALIDEZ

Empezemos por el caso más sencillo: una prueba cuyos resultados posibles son "positivo" o "negativo". Para evaluar su validez habrá que aplicarla a una muestra de individuos que sepamos que tienen la enfermedad y a otra que sepamos que no la tiene. Los resultados se pueden expresar en una tabla como la siguiente:

Resultado de la Prueba	Estado de paciente		Total
	No Enfermo	Enfermo	
Negativo	a	b	r
Positivo	c	d	s
Total	t	u	N

siendo N el número total de individuos observados, t el número de no enfermos, u el de enfermos, a el número de no enfermos en que la prueba ha dado negativa, etc.

Si la prueba fuera perfectamente válida $b=c=0$. Como índices de validez se definen:

- Sensibilidad (S)** o proporción de verdaderos positivos como la probabilidad de que la prueba dé positivo condicionada a que el individuo esté enfermo.
- Especificidad (E)** o proporción de verdaderos negativos como la probabilidad de que la prueba dé negativo condicionada a que el individuo no esté enfermo.
- También se usan como índices de error sus complementarios, es decir **proporción de falsos positivos (PFP)** y de **falsos negativos (PFN)**.

A partir de la tabla, estos índices (que tomarán valores entre 0 y 1) se estiman como:

- $S = d/u$
- $E = a/t$
- $PFP = c/t = 1-E$, y $PFN = b/u = 1-S$

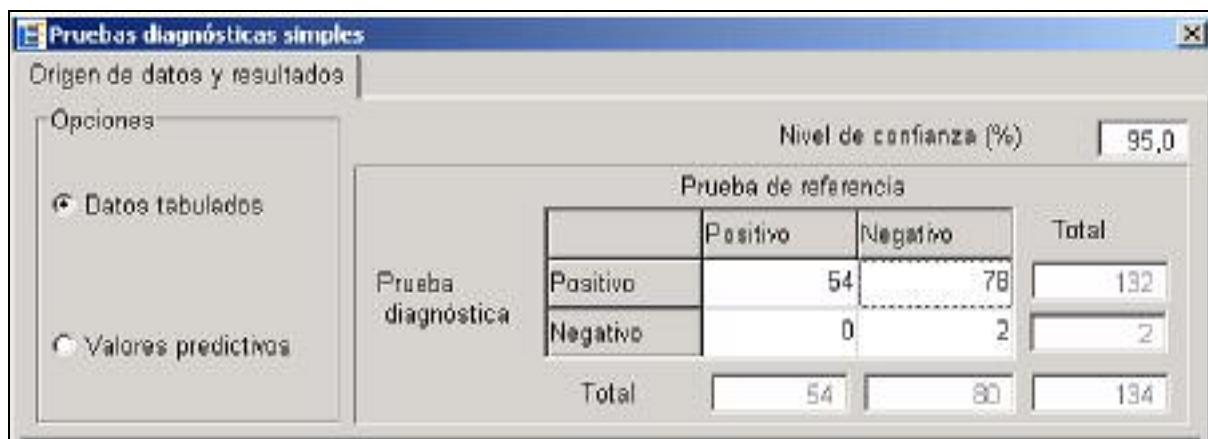
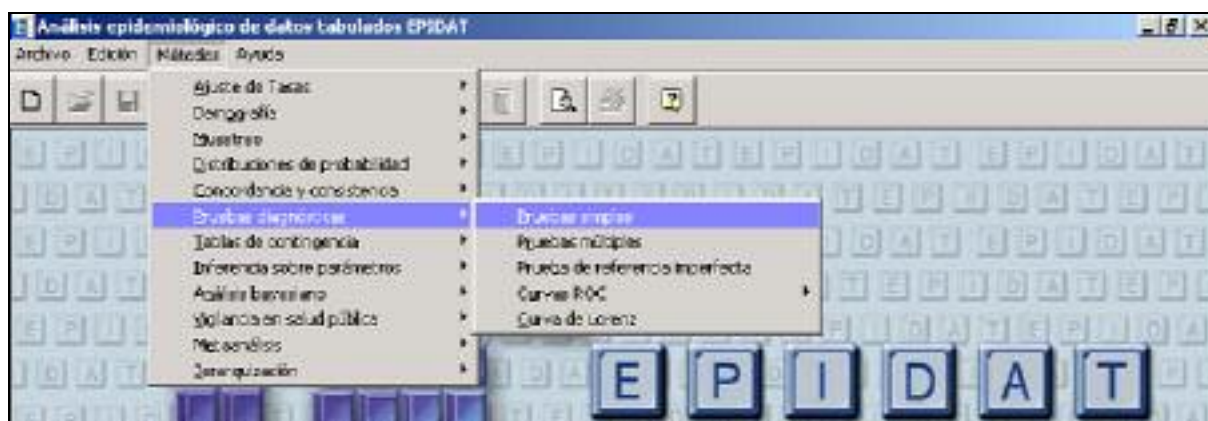
Como son proporciones, sus intervalos de confianza se construyen como tal, es decir, por ejemplo para la sensibilidad y asumiendo muestras grandes, su intervalo de confianza del 95% sería:

$$I C .100(1-\alpha): \hat{S} \pm z_{\alpha/2} \sqrt{\frac{\hat{S}(1-\hat{S})}{n}}$$

Ejemplo 1: Evaluación de la de perfusión en el diagnóstico de embolismo pulmonar, donde se tomó como prueba de referencia (o *gold standard*) la arteriografía pulmonar:

Prueba diagnóstica (gammagrafía)	Estado del paciente (arteriografía pulmonar)	
	Enfermo	No Enfermo
Positivo	54	78
Negativo	0	2

Con **EpiDat** obtendríamos,



[1] Pruebas diagnósticas simples			
Nivel de confianza:		95,0%	
Prueba diagnóstica	Prueba de referencia		Total
	Enfermos	Sanos	
Positivo	54	78	132
Negativo	0	2	2
Total	54	80	134
	Valor	IC (95%)	
Sensibilidad (%)	100,00	99,07	100,00
Especificidad (%)	2,50	1,85	3,15
Índice de Validez (%)	41,79	41,37	42,21
Valor predictivo + (%)	40,91	40,48	41,34
Valor predictivo - (%)	100,00	75,00	100,00
Prevalencia (%)	40,30	39,87	40,72
Índice de Youden	0,03	0,02	0,03
Razón de verosimilitud +	1,03	1,03	1,03
Razón de verosimilitud -	-	-	-

Es una prueba extraordinariamente sensible (en todos los enfermos da positivo, S=1) pero muy inespecífica (también da positivo en muchos individuos no enfermos, E=0,025).

2. VALORES PREDICTIVOS

A partir del ejemplo anterior, podemos plantearnos si sería razonable usar esa prueba, dado que únicamente permite clasificar correctamente individuos enfermos, pero no así con los no enfermos. Para ello Hay otros índices que ayudan en esta decisión, son los llamados **valores predictivos**:

- Valor Predictivo Positivo (PPP):** proporción de individuos con resultado “positivo” que verdaderamente tienen la enfermedad.
- Valor Predictivo Negativo (PPN):** proporción de individuos con resultado “negativo” que verdaderamente no tienen la enfermedad.

A partir de la tabla, los valores predictivos (que tomarán valores entre 0 y 1) se estiman como:

- $VP+ = P(\text{enfermo}|\text{prueba positiva}) = d/s$
- $VP- = P(\text{no enfermo}|\text{prueba negativa}) = a/r$

Para el ejemplo anterior, tenemos que $VPP = 54/132 = 0,4091$ y $VPN = 2/2 = 1$, que ponen de manifiesto que podría ser una muy buena prueba para descartar un diagnóstico del embolismo pulmonar, pero mala para confirmarlo.

Hay que tener en cuenta que los valores predictivos se corresponden a probabilidades *postprueba* (probabilidades condicionales) donde la probabilidad *preprueba*, estimada a partir de la prevalencia de la enfermedad en el estudio, en el ejemplo anterior sería: $54/134 = 0,40$. Por este motivo, los valores predictivos no serían aplicables a otras situaciones. Sin embargo, el **Teorema de Bayes** permite calcular los valores predictivos a partir de la sensibilidad y especificidad (que sólo dependen de los valores de la prueba) para distintas probabilidades preprueba o prevalencia:

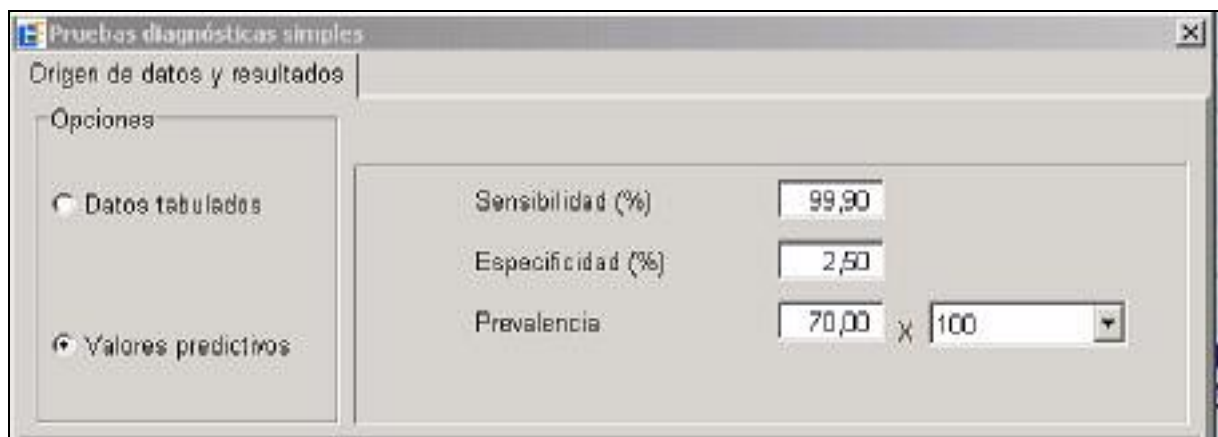
$$VP+ = p(E | P+) = \frac{p(P+ | E)p(E)}{p(P+ | E)p(E) + p(P+ | \bar{E})p(\bar{E})} = \frac{Sen \cdot p(E)}{Sen \cdot p(E) + (1 - Esp)(1 - p(E))}$$

siendo p(E) la probabilidad preprueba y similarmente:

$$VP- = \frac{Esp(1 - p(E))}{Esp(1 - p(E)) + (1 - Sen)p(E)}$$

Ejemplo 2: Calcular los VPP y VPN de la gammagrafía de perfusión para un paciente cuya probabilidad preprueba es de 0,7

Con **EpiDat** obtenemos,



```
[ 2] Pruebas diagnósticas simples

      Sensibilidad:      99,90%
      Especificidad:    2,50%
      Prevalencia:      70,00 x      100

-----
      Valor
-----
      Índice de Validez (%)      70,68
      Valor predictivo + (%)     70,51
      Valor predictivo - (%)     91,46

      Índice de Youden           0,02
      Razón de verosimilitud +  1,02
      Razón de verosimilitud -  0,04
```

Obsérvese que ahora el VPP ha mejorado sustancialmente, mientras que el VPN se mantiene prácticamente constante.

3. CORRECCIÓN DEL CÁLCULO DE LA RAZÓN DE ODDS

Si en un estudio de casos y controles tenemos que el diagnóstico de la exposición al factor en estudio se realiza con una prueba cuya **sensibilidad** $\alpha = \Pr(T+|E+)$ y **especificidad** $\beta = \Pr(T-|E-)$, debemos analizar cuanto influyen los errores de clasificación en los resultados obtenidos.

En las siguientes tablas representamos los resultados del estudio de caso control con clasificación correcta y con clasificación por el método anterior (sesgo de medición).

Exposición	Sin sesgo de medición		Con sesgo de medición	
	Caso	Control	Caso	Control
Expuesto	a	b	a'	b'
No Expuesto	c	d	c'	d'
Total	n_1	n_0	n_1	n_0

Sabemos que $a + c = a' + c' = n_1$, $b + d = b' + d' = n_0$ y considerando la sensibilidad y especificidad del método de diagnóstico de la exposición obtenemos las siguientes relaciones:

$$\text{Casos: } \begin{cases} a' = a\alpha + c(1 - \beta) \\ c' = a(1 - \alpha) + c\beta \end{cases} \quad \text{Controles: } \begin{cases} b' = b\alpha + d(1 - \beta) \\ d' = b(1 - \alpha) + d\beta \end{cases}$$

Ejemplo 3: Un investigador obtuvo los siguientes resultados al comparar respuestas obtenidas personalmente y por parientes de los individuos en estudio, en cuanto al hábito de fumar.

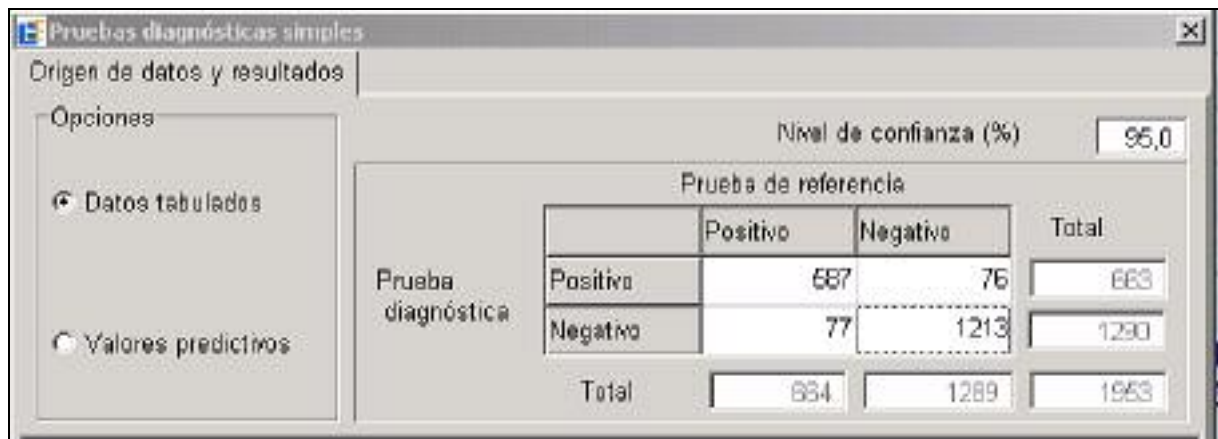
Respuesta por el pariente	Respuesta por el individuo en estudio	
	Fuma	No fuma
Fuma	587	76
No Fuma	77	1213

Después condujo una investigación de casos y controles, sobre cáncer de pulmón y hábito de fumar, con respuestas dadas por los parientes. Los resultados se muestran en la siguiente tabla.

	Casos	Controles
Fumador	135	282
No fumador	65	328

De donde obtenemos $OR = 2.42$ con un intervalo de confianza del 95%: (1.73, 3.38). Sin embargo, sabemos que el método “información dada por pariente” para diagnosticar la exposición, tiene una sensibilidad $\alpha = 587/664 = 0.88$ y una especificidad $\beta = 1213/1289 = 0.94$.

Con EpiDat obtendríamos,



[1] Pruebas diagnósticas simples

Nivel de confianza: 95,0%

Prueba diagnóstica	Prueba de referencia		Total
	Enfermos	Sanos	
Positivo	587	76	663
Negativo	77	1213	1290
Total	664	1289	1953

	Valor	IC (95%)	
Sensibilidad (%)	88,40	88,31	88,49
Especificidad (%)	94,10	94,06	94,15
Índice de Validez (%)	92,17	92,13	92,20
Valor predictivo + (%)	88,54	88,45	88,63
Valor predictivo - (%)	94,03	93,98	94,08
Prevalencia (%)	34,00	33,96	34,04
Índice de Youden	0,83	0,82	0,83
Razón de verosimilitud +	14,99	14,97	15,01
Razón de verosimilitud -	0,12	0,12	0,12

Así, utilizando los sistemas de ecuaciones anteriores obtenemos:

$$\text{Casos: } \begin{cases} 135 = a \times 0.88 + c \times 0.06 \\ 65 = a \times 0.12 + c \times 0.94 \end{cases}, \text{ por tanto } a = 150, \text{ y } c = 50.$$

$$\text{Controles: } \begin{cases} 282 = a \times 0.88 + c \times 0.06 \\ 328 = a \times 0.12 + c \times 0.94 \end{cases}, \text{ por tanto } b = 300, \text{ y } d = 310.$$

El verdadero OR sería igual a 3.1 con un intervalo de confianza del 95%: (2.17, 4.43), por tanto el error de clasificación sesga el resultado hacia la hipótesis nula.

4. CORRECCIÓN DEL CÁLCULO DEL RIESGO RELATIVO

Si en un estudio de cohortes tenemos que el diagnóstico de la enfermedad se realiza con una prueba cuya sensibilidad $\alpha = \Pr(\mathbf{T+|D+})$ y **sensibilidad** $\beta = \Pr(\mathbf{T-|D-})$, de nuevo, debemos analizar cuanto influyen los errores de clasificación en los resultados obtenidos.

En las siguientes tablas representamos los resultados del estudio de cohorte con clasificación correcta y con clasificación por el método anterior (sesgo de medición).

	Sin error de clasificación		Con error de clasificación		Total
	D+	D-	D+	D-	
E+	a	b	a'	b'	m ₁
E-	c	d	c'	d'	m ₀

Sabemos que $a + b = a' + b' = m_1$, $c + d = c' + d' = m_0$ y considerando la sensibilidad y especificidad del método de diagnóstico obtenemos las siguientes relaciones:

$$\text{Expuestos: } \begin{cases} a' = a\alpha + b(1 - \beta) \\ b' = a(1 - \alpha) + b\beta \end{cases} \quad \text{No Expuestos: } \begin{cases} c' = c\alpha + d(1 - \beta) \\ d' = c(1 - \alpha) + d\beta \end{cases}$$

Ejemplo 4: En un estudio de seguimiento de 1000 mujeres, 500 expuestas a las sustancias químicas que se emplean en los salones de belleza, y 500 no expuestas, se detectó que 100 mujeres de la cohorte expuesta y 50 de la cohorte no expuesta tuvieron aborto espontáneo. Los resultados se presentan en la siguiente tabla:

	D+	D-	Total
E+	100	400	500
E-	50	450	500

De donde obtenemos $RR = 2.0$ y un intervalo de confianza del 95%: (1.46, 2.74). Supongamos que en este estudio se utiliza una prueba cuya sensibilidad es 90% y especificidad 90%, ¿cuales serían los resultados entonces? Aplicando las fórmulas anteriores obtenemos.

	En Expuestos E+			En No Expuestos E-		
	D+	D-	Total	D+	D-	Total
D+	90	40	130	45	45	90
D-	10	360	370	5	405	410

El RR sería igual a 1.44 y un intervalo de confianza del 95%: (1.14, 1.83), por tanto el error de clasificación sesga el resultado hacia la hipótesis nula.